

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Yang Wang, Michael W. Spellman

(ii) TITLE OF INVENTION: O-Fucosyltransferase

(iii) NUMBER OF SEQUENCES: 17

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Genentech, Inc.
- (B) STREET: 1 DNA Way
- (C) CITY: South San Francisco
- (D) STATE: California
- (E) COUNTRY: USA
- (F) ZIP: 94080

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: WinPatin (Genentech)

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: Unassigned
- (B) FILING DATE: 26-Nov-1997
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: 08/792498
- (B) FILING DATE: 31 January 1997

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Svoboda, Craig G.
- (B) REGISTRATION NUMBER: 39,044
- (C) REFERENCE/DOCKET NUMBER: P1041P1

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 650/225-1489
- (B) TELEFAX: 650/952-9881

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1514 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGCCCGCGG GCTCCTGGGA CCCGGCCGGT TACCTGCTCT ACTGCCCTG 50
CATGGGGCGC TTTGGGAACC AGGCCGATCA CTTCTGGGC TCTCTGGCAT 100
TTGCAAAGCT GCTAAACCGT ACCTTGGCTG TCCCTCCTTG GATTGAGTAC 150
CAGCATCACA AGCCTCCTT CACCAACCTC CATGTGTCCT ACCAGAAGTA 200
CTTCAAGCTG GAGCCCCTCC AGGCTTACCA TCGGGTCATC AGCTTGGAGG 250
ATTTCATGGA GAAGCTGGCA CCCACCCACT GGCCCCCTGA GAAGCGGGTG 300
GCATACTGCT TTGAGGTGGC AGCCCAGCGA AGCCCAGATA AGAAGACGTG 350
CCCCATGAAG GAAGGAAACC CCTTTGGCCC ATTCTGGGAT CAGTTCATG 400
TGAGTTCAA CAAGTCGGAG CTTTTACAG GCATTTCCTT CAGTGCTTCC 450
TACAGAGAAC AATGGAGCCA GAGATTTCT CCAAAGGAAC ATCCGGTGCT 500
TGCCCTGCCA GGAGCCCCAG CCCAGTTCCC CGTCCTAGAA GAACACAGGC 550
CACTACAGAA GTACATGGTA TGGTCAGACG AAATGGTGAA GACGGGAGAG 600
GCCCAGATTG ATGCCCACCT TGTCCGGCCC TATGTGGCA TTCATCTGCG 650
CATTGGCTCT GACTGGAAGA ACGCCTGTGC CATGCTGAAG GACGGGACTG 700
CAGGCTCGCA CTTCATGGCC TCTCCGCAGT GTGTGGGCTA CAGCCGCAGC 750
ACAGCGGCCG CCCTCACGAT GACTATGTGC CTGCCTGACC TGAAGGAGAT 800
CCAGAGGGCT GTGAAGCTCT GGGTGAGGTC GCTGGATGCC CAGTCGGTCT 850

ACGTTGCTAC TGATTCCGAG AGTTATGTGC CTGAGCTCCA ACAGCTCTTC 900
AAAGGGAAGG TGAAGGTGGT GAGCCTGAAG CCTGAGGTGG CCCAGGTCGA 950
CCTGTACATC CTCGGCCAAG CCGACCAC TTATTGGCAAC TGTGTCTCCT 1000
CCTTCACTGC CTTTGTGAAG CGGGAGCGGG ACCTCCAGGG GAGGCCGTCT 1050
TCTTTCTTCG GCATGGACAG GCCCCCTAAG CTGCGGGACG AGTTCTGATT 1100
CTGGCCGGAG CACCAGACCC TCTGATCCTG GAGGGACCAG AGTCTGAGCT 1150
GGTCCTTCCA GCCAGGCCTG GCAGCCAGAG GTGCTCCGGG ATTGCAAAC 1200
CCTCTTCTCA CCTGCCAAAG ATGGAGAAGA GTGCCAGGG CCCCTCAAGG 1250
AGGGAGACGC TCCATATCCC AGGGCATAGG ACTTGCAGGT TCCTAGGAGC 1300
AGGAGCATCT CCCATCGCAC GTGCTTCTG CTCTTCTGGG AATTCTCAC 1350
ACTGGCAAAG CAGTCCAGCC TCCGTCTTCT GGTCCACTCT GCTCTGAGCA 1400
GCCTGGGATG CTGAACCTTT CAGAGAGATT TTTTATAGA GAGATTCTA 1450
TAATTTGAT ACAAGGTCA GACTATCCTA GAACTCTCTG TGGTTTTGA 1500
AAATCATTGA ATTC 1514

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 365 amino acids
(B) TYPE: Amino Acid
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Pro	Ala	Gly	Ser	Trp	Asp	Pro	Ala	Gly	Tyr	Leu	Leu	Tyr	Cys
1					5					10				15
Pro	Cys	Met	Gly	Arg	Phe	Gly	Asn	Gln	Ala	Asp	His	Phe	Leu	Gly
					20					25				30

Ser	Leu	Ala	Phe	Ala	Lys	Leu	Leu	Asn	Arg	Thr	Leu	Ala	Val	Pro
						35					40			45
Pro	Trp	Ile	Glu	Tyr	Gln	His	His	Lys	Pro	Pro	Phe	Thr	Asn	Leu
						50				55			60	
His	Val	Ser	Tyr	Gln	Lys	Tyr	Phe	Lys	Leu	Glu	Pro	Leu	Gln	Ala
					65				70				75	
Tyr	His	Arg	Val	Ile	Ser	Leu	Glu	Asp	Phe	Met	Glu	Lys	Leu	Ala
					80				85				90	
Pro	Thr	His	Trp	Pro	Pro	Glu	Lys	Arg	Val	Ala	Tyr	Cys	Phe	Glu
					95				100				105	
Val	Ala	Ala	Gln	Arg	Ser	Pro	Asp	Lys	Lys	Thr	Cys	Pro	Met	Lys
					110				115				120	
Glu	Gly	Asn	Pro	Phe	Gly	Pro	Phe	Trp	Asp	Gln	Phe	His	Val	Ser
					125				130				135	
Phe	Asn	Lys	Ser	Glu	Leu	Phe	Thr	Gly	Ile	Ser	Phe	Ser	Ala	Ser
					140				145				150	
Tyr	Arg	Glu	Gln	Trp	Ser	Gln	Arg	Phe	Ser	Pro	Lys	Glu	His	Pro
					155				160				165	
Val	Leu	Ala	Leu	Pro	Gly	Ala	Pro	Ala	Gln	Phe	Pro	Val	Leu	Glu
					170				175				180	
Glu	His	Arg	Pro	Leu	Gln	Lys	Tyr	Met	Val	Trp	Ser	Asp	Glu	Met
					185				190				195	
Val	Lys	Thr	Gly	Glu	Ala	Gln	Ile	His	Ala	His	Leu	Val	Arg	Pro
					200				205				210	
Tyr	Val	Gly	Ile	His	Leu	Arg	Ile	Gly	Ser	Asp	Trp	Lys	Asn	Ala
					215				220				225	
Cys	Ala	Met	Leu	Lys	Asp	Gly	Thr	Ala	Gly	Ser	His	Phe	Met	Ala
					230				235				240	

Ser	Pro	Gln	Cys	Val	Gly	Tyr	Ser	Arg	Ser	Thr	Ala	Ala	Pro	Leu	
													245	250	255
Thr	Met	Thr	Met	Cys	Leu	Pro	Asp	Leu	Lys	Glu	Ile	Gln	Arg	Ala	
													260	265	270
Val	Lys	Leu	Trp	Val	Arg	Ser	Leu	Asp	Ala	Gln	Ser	Val	Tyr	Val	
													275	280	285
Ala	Thr	Asp	Ser	Glu	Ser	Tyr	Val	Pro	Glu	Leu	Gln	Gln	Leu	Phe	
													290	295	300
Lys	Gly	Lys	Val	Lys	Val	Val	Ser	Leu	Lys	Pro	Glu	Val	Ala	Gln	
													305	310	315
Val	Asp	Leu	Tyr	Ile	Leu	Gly	Gln	Ala	Asp	His	Phe	Ile	Gly	Asn	
													320	325	330
Cys	Val	Ser	Ser	Phe	Thr	Ala	Phe	Val	Lys	Arg	Glu	Arg	Asp	Leu	
													335	340	345
Gln	Gly	Arg	Pro	Ser	Ser	Phe	Phe	Gly	Met	Asp	Arg	Pro	Pro	Lys	
													350	355	360
Leu	Arg	Asp	Glu	Phe											
														365	

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 61 amino acids
 - (B) TYPE: Amino Acid
 - (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Arg	Leu	Ala	Gly	Ser	Trp	Asp	Leu	Ala	Gly	Tyr	Leu	Leu	Tyr	Xaa			
														1	5	10	15
Pro	Xaa	Met	Gly	Arg	Phe	Gly	Asn	Gln	Ala	Asp	His	Phe	Leu	Gly			
														20	25	30	

Ser Leu Ala Phe Ala Lys Leu Xaa Val Arg Thr Leu Ala Val Pro
35 40 45

Pro Trp Ile Glu Tyr Gln His His Lys Pro Pro Phe Thr Asn Leu
50 55 60

His
61

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1300 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TTATTCATAC CGTCCCACCA TCGGGCGCGG ATCAGATCCA TGGCCAAGTT 50
CCTGGTCAAC GTGGCCCTGC TGCTGCTGCT GCTGCTGCTG TCCGGAGCCT 100
GGGCCCATAT GAGATCCCCT CACCACATCACC ATCACATGCC CGCGGGCTCC 150
TGGGACCCGG CCGGTTACCT GCTCTACTGC CCCTGCATGG GGCGCTTTGG 200
GAACCAGGCC GATCACTTCT TGGGCTCTCT GGCATTTGCA AAGCTGCTAA 250
ACCGTACCTT GGCTGTCCCT CCTTGGATTG AGTACCAGCA TCACAAGCCT 300
CCTTTCACCA ACCTCCATGT GTCCTACCAG AAGTACTTCA AGCTGGAGCC 350
CCTCCAGGCT TACCATCGGG TCATCAGCTT GGAGGATTTC ATGGAGAAGC 400
TGGCACCCAC CCACTGGCCC CCTGAGAAGC GGGTGGCATA CTGCTTTGAG 450
GTGGCAGGCC AGCGAAGCCC AGATAAGAAG ACGTGCCCCA TGAAGGAAGG 500
AAACCCCTT GGCCCATTCT GGGATCAGTT TCATGTGAGT TTCAACAAGT 550
CGGAGCTTT TACAGGCATT TCCCTCAGTG CTTCCCTACAG AGAACAAATGG 600

AGCCAGAGAT TTTCTCCAAA GGAACATCCG GTGCTTGCCT TGCCAGGAGC 650
CCCAGCCAG TTCCCCGTCC TAGAGGAACA CAGGCCACTA CAGAAGTACA 700
TGGTATGGTC AGACGAAATG GTGAAGACGG GAGAGGCCA GATTCATGCC 750
CACCTTGTCC GGCCCTATGT GGGCATTCA CTGCGCATTG GCTCTGACTG 800
GAAGAACGCC TGTGCCATGC TGAAGGACGG GACTGCAGGC TCGCACTTCA 850
TGGCCTCTCC GCAGTGTGTG GGCTACAGCC GCAGCACAGC GGCCCCCTC 900
ACGATGACTA TGTGCCTGCC TGACCTGAAG GAGATCCAGA GGGCTGTGAA 950
GCTCTGGGTG AGGTCGCTGG ATGCCAGTC GGTCTACGTT GCTACTGATT 1000
CCGAGAGTTA TGTGCCTGAG CTCCAACAGC TCTTCAAAGG GAAGGTGAAG 1050
GTGGTGAGCC TGAAGCCTGA GGTGGCCAG GTGACCTGT ACATCCTCGG 1100
CCAAGCCGAC CACTTATTG GCAACTGTGT CTCCTCCTTC ACTGCCTTG 1150
TGAAGCGGGA GCAGGACCTC CAGGGGAGGC CGTCTCTTT CTTCGGCATG 1200
GACAGGCCCT CTAAGCTGCG GGACGAGTTC TGATTCTGGC CGGAGCACCA 1250
GACCCTCTGA TCCTGGAGGG ACCAGAGTCT GAGCTGGTCC TTCCAGCCAG 1300

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11284 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAGCTTTACT CGTAAAGCGA GTTGAAGGAT CATATTTAGT TGCCTTATG 50
AGATAAGATT GAAAGCACGT GTAAAATGTT TCCCGCGCGT TGGCACAACT 100

ATTTACAATG CGGCCAAGTT ATAAAAGATT CTAATCTGAT ATGTTTAAA 150
ACACCTTGCGTAC GTGACTAGCG AAGAAGATGT 200
GTGGACCGCA GAACAGATAG TAAAACAAAA CCCTAGTATT GGAGCAATAA 250
TCGATTTAAC CAACACGTCT AAATATTATG ATGGTGTGCA TTTTTGCGG 300
GCGGGCCTGT TATACAAAAA AATTCAAGTA CCTGGCCAGA CTTTGCCGCC 350
TGAAAGCATA GTTCAAGAAT TTATTGACAC GGTAAAAGAA TTTACAGAAA 400
AGTGTCCCGG CATGTTGGTG GGCGTGCCT GCACACACGG TATTAATCGC 450
ACCGGTTACA TGGTGTGCAG ATATTTAACATG CACACCTGG GTATTGCGCC 500
GCAGGAAGCC ATAGATAGAT TCGAAAAAGC CAGAGGTCAC AAAATTGAAA 550
GACAAAATTA CGTTCAAGAT TTATTAATTT AATTAATATT ATTTGCATTC 600
TTAACAAAT ACTTTATCCT ATTTCAAAT TGTTGCGCTT CTTCCAGCGA 650
ACCAAAACTA TGCTTCGCTT GCTCCGTTA GCTTGTAGCC GATCAGTGGC 700
GTTGTTCAA TCGACGGTAG GATTAGGCCG GATATTCTCC ACCACAATGT 750
TGGCAACGTT GATGTTACGT TTATGCTTT GGTTTCCAC GTACGTCTTT 800
TGGCCGGTAA TAGCCGTAAA CGTAGTGCCG TCGCGCGTCA CGCACAAACAC 850
CGGATGTTTG CGCTTGTCCG CGGGGTATTG AACCGCGCGA TCCGACAAAT 900
CCACCACTTT GGCAACTAAA TCGGTGACCT GCGCGTCTT TTTCTGCATT 950
ATTCGTCTT TCTTTGCAT GGTTCCCTGG AAGCCGGTGT ACATGCGGTT 1000
TAGATCAGTC ATGACGCGCG TGACCTGCAA ATCTTGGCC TCGATCTGCT 1050
TGTCTTGAT GGCAACGATG CGTTCAATAA ACTCTTGTCTT TTTAACAAAGT 1100
TCCTCGGTTT TTTGCGGCCAC CACCGCTTGC AGCGCGTTG TGTGCTCGGT 1150

GAATGTCGCA ATCAGCTTAG TCACCAACTG TTTGCTCTCC TCCTCCCGTT 1200
GTTTGATCGC GGGATCGTAC TTGCCGGTGC AGAGCACTTG AGGAATTACT 1250
TCTTCTAAAA GCCATTCTTG TAATTCTATG GCGTAAGGCA ATTTGGACTT 1300
CATAATCAGC TGAATCACGC CGGATTTAGT AATGAGCACT GTATGCGGCT 1350
GCAAATACAG CGGGTCGCCCG CTTTCACGA CGCTGTTAGA GGTAGGGCCC 1400
CCATTTGGA TGGTCTGCTC AAATAACGAT TTGTATTTAT TGTCTACATG 1450
AACACGTATA GCTTTATCAC AAACGTATA TTTTAAACTG TTAGCGACGT 1500
CCTTGGCCAC GAACCGGACC TGTTGGTCGC GCTCTAGCAC GTACCGCAGG 1550
TTGAACGTAT CTTCTCCAAA TTTAAATTCT CCAATTAA CGCGAGCCAT 1600
TTTGATACAC GTGTGTCGAT TTTGCAACAA CTATTGTTT TTAACGCAA 1650
CTAAACTTAT TGTGGTAAGC AATAATTAAA TATGGGGAA CATGCGCCGC 1700
TACAACACTC GTCGTTATGA ACGCAGACGG CGCCGGTCTC GGCAGCAAGCG 1750
GCTAAAACGT GTTGCACGTT CAACGCGCA AACATCGCAA AAGCCAATAG 1800
TACAGTTTG ATTTGCATAT TAACGGCGAT TTTTAAATT ATCTTATTAA 1850
ATAAAATAGTT ATGACGCCTA CAACTCCCCG CCCGCGTTGA CTCGCTGCAC 1900
CTCGAGCAGT TCGTTGACGC CTTCCCTCCGT GTGGCCGAAC ACGTCGAGCG 1950
GGTGGTCGAT GACCAGCGGC GTGCCGCACCG CGACGCACAA GTATCTGTAC 2000
ACCGAATGAT CGTCGGCGA AGGCACGTG GCCTCCAAGT GGCAATATTG 2050
GCAAATTCGA AAATATATAC AGTTGGTTG TTTGCGCATA TCTATCGTGG 2100
CGTTGGGCAT GTACGTCCGA ACGTTGATT GCATGCAAGC CGAAATTAAA 2150
TCATTGCGAT TAGTGCAGT AAAACGTTGT ACATCCTCGC TTTTAATCAT 2200
GCCGTCGATT AAATCGCGCA ATCGAGTCAG GTGATCAAAG TGTGGAATAA 2250

TGTTTTCTTT GTATTCCCGA GTCAAGCGCA GCGCGTATTT TAACAAACTA 2300
GCCATCTTGT AAGTTAGTTT CATTAAATGC AACTTTATCC AATAATATAT 2350
TATGTATCGC ACGTCAAGAA TTAACAATGC GCCCGTTGTC GCATCTCAAC 2400
ACGACTATGA TAGAGATCAA ATAAAGCGCG AATTAAATAG CTTGCGACGC 2450
AACGTGCACG ATCTGTGCAC GCGTTCCGGC ACGAGCTTG ATTGTAATAA 2500
GTTTTACGA AGCGATGACA TGACCCCCGT AGTGACAACG ATCACGCCA 2550
AAAGAACTGC CGACTACAAA ATTACCGAGT ATGTCGGTGA CGTTAAAACT 2600
ATTAAGCCAT CCAATCGACC GTTAGTCGAA TCAGGACCGC TGGTGCAGA 2650
AGCCGCGAAG TATGGCGAAT GCATCGTATA ACGTGTGGAG TCCGCTCATT 2700
AGAGCGTCAT GTTGTAGACAA GAAAGCTACA TATTTAATTG ATCCCGATGA 2750
TTTATTGAT AAATTGACCC TAACTCCATA CACGGTATTC TACAATGGCG 2800
GGGTTTTGGT CAAAATTCC GGACTGCGAT TGTACATGCT GTTAACGGCT 2850
CCGCCCACTA TTAATGAAAT TAAAAATTCC AATTTAAAAA AACGCAGCAA 2900
GAGAAACATT TGTATGAAAG AATGCGTAGA AGGAAAGAAA AATGTCGTCG 2950
ACATGCTGAA CAACAAGATT AATATGCCTC CGTGTATAAA AAAAATATTG 3000
AACGATTGAA AAGAAAACAA TGTACCGCGC GGCGGTATGT ACAGGAAGAG 3050
GTTTATACTA AACTGTTACA TTGCAAACGT GGTTCTGTGT GCCAAGTGTG 3100
AAAACCGATG TTTAATCAAG GCTCTGACGC ATTTCTACAA CCACGACTCC 3150
AAAGTGTGTGG GTGAAGTCAT GCATCTTTA ATCAAATCCC AAGATGTGTA 3200
TAAACCACCA AACTGCCAAA AAATGAAAAC TGTCGACAAG CTCTGTCCGT 3250
TTGCTGGCAA CTGCAAGGGT CTCAATCCTA TTTGTAATTA TTGAATAATA 3300

AAACAATTAT AAATGCTAAA TTTGTTTTT ATTAAACGATA CAAACCAAAC 3350
GCAACAAGAA CATTGTAGT ATTATCTATA ATTGAAAACG CGTAGTTATA 3400
ATCGCTGAGG TAATATTTAA AATCATTTTC AAATGATTCA CAGTTAATT 3450
GCGACAATAT AATTTTATTTC TCACATAAAC TAGACGCCTT GTCGTCTTCT 3500
TCTTCGTATT CCTTCTCTTT TTCATTTTC TCCTCATAAA AATTAACATA 3550
GTTATTATCG TATCCATATA TGTATCTATC GTATAGAGTA AATTTTTGT 3600
TGTCTAAAT ATATATGTCT TTTTAATGG GGTGTATAGT ACCGCTGCGC 3650
ATAGTTTTTC TGTAATTAC AACAGTGCTA TTTTCTGGTA GTTCTTCGGA 3700
GTGTGTTGCT TTAATTATTA AATTATATA ATCAATGAAT TTGGGATCGT 3750
CGGTTTGTA CAATATGTTG CCGGCATAGT ACGCAGCTTC TTCTAGTTCA 3800
ATTACACCAC TTTTAGCAG CACCGGATTA ACATAACTTT CCAAAATGTT 3850
GTACGAACCG TTAAACAAAA ACAGTTCACC TCCCTTTCT ATACTATTGT 3900
CTGCGAGCAG TTGTTGTTG TTAAAATAA CAGCCATTGT AATGAGACGC 3950
ACAAACTAAT ATCACAAACT GGAAATGTCT ATCAATATAT AGTGCTGAT 4000
ATCATGGAGA TAATTAAAAT GATAACCATC TCGCAAATAA ATAAGTATT 4050
TACTGTTTC GTAACAGTT TGTAATAAAA AAACCTATAA ATATTCCGGA 4100
TTATTCTACAC CGTCCCACCA TCAGGGCGCGG ATCAGATCCA TGGCCAAGTT 4150
CCTGGTCAAC GTGGCCCTGC TGCTGCTGCT GCTGCTGCTG TCCGGAGCCT 4200
GGGCCATAT GAGATCCCAC CACCATCACC ATCACATGCC CGCGGGCTCC 4250
TGGGACCCGG CCGGTTACCT GCTCTACTGC CCCTGCATGG GGCGCTTGG 4300
GAACCAGGCC GATCACTTCT TGGGCTCTCT GGCATTTGCA AAGCTGCTAA 4350
ACCGTACCTT GGCTGTCCCT CCTGGATTG AGTACCAAGCA TCACAAGCCT 4400

CCTTCACCA ACCTCCATGT GTCCTACCAAG AAGTACTTCA AGCTGGAGCC 4450
CCTCCAGGCT TACCATCGGG TCATCAGCTT GGAGGATTTC ATGGAGAAGC 4500
TGGCACCCAC CCACCTGGCCC CCTGAGAAGC GGGTGGCATA CTGCTTGAG 4550
GTGGCAGCCC AGCGAAGCCC AGATAAGAAG ACGTCCCCA TGAAGGAAGG 4600
AAACCCCTT GGCCCATTCT GGGATCAGTT TCATGTGAGT TTCAACAAGT 4650
CGGAGCTTT TACAGGCATT TCCTTCAGTG CTTCTACAG AGAACAAATGG 4700
AGCCAGAGAT TTTCTCCAAA GGAACATCCG GTGCTTGCCC TGCCAGGAGC 4750
CCCAGCCCAG TTCCCCGTCC TAGAGGAACA CAGGCCACTA CAGAAGTACA 4800
TGGTATGGTC AGACGAAATG GTGAAGACGG GAGAGGCCA GATTCAATGCC 4850
CACCTTGTCC GGCCCTATGT GGGCATTCACT CTGCGCATTG GCTCTGACTG 4900
GAAGAACGCC TGTGCCATGC TGAAGGACGG GACTGCAGGC TCGCACTTCA 4950
TGGCCTCTCC GCAGTGTGTG GGCTACAGCC GCAGCACAGC GGCCCCCTC 5000
ACGATGACTA TGTGCCTGCC TGACCTGAAG GAGATCCAGA GGGCTGTGAA 5050
GCTCTGGGTG AGGTCGCTGG ATGCCAGTC GGTCTACGTT GCTACTGATT 5100
CCGAGAGTTA TGTGCCTGAG CTCCAACAGC TCTTCAAAGG GAAGGTGAAG 5150
GTGGTGAGCC TGAAGCCTGA GGTGGCCAG GTCGACCTGT ACATCCTCGG 5200
CCAAGCCGAC CACTTTATTG GCAACTGTGT CTCCTCCTTC ACTGCCTTG 5250
TGAAGCGGGA GCGGGACCTC CAGGGGAGGC CGTCTTCTT CTTCGGCATG 5300
GACAGGCCAG CTAAGCTGCG GGACGAGTTC TGATTCTGGC CGGAGCACCA 5350
GACCCTCTGA TCCTGGAGGG ACCAGAGTCT GAGCTGGTCC TTCCAGGCCAG 5400
GCCTGGCAGC CAGAGGTGCT CCGGGATTGC AAACCTCCTCT TCTCACCTGC 5450

CAAAGATGGA GAAGAGTGCC AGGGACCCCT CAAGGAGGGA GACGCTCCAT 5500
ATCCCAGGGC ATAGGACTTG CAGGTTCCCTA GGAGCAGGAG CATCTCCCAT 5550
CGCACGTGCT TTCTGCTCTT CTGGGAATT CTCACACTGG CAAAGCAGTC 5600
CAGCCTCCGT CTTCTGGTCC ACTCTGCTCT GAGCAGCCTG GGATGCTGAA 5650
CTCTTCAGAG AGATTTTTT ATAGAGAGAT TTCTATAATT TTGATACAAG 5700
GTCATGACTA TCCTAGAACT CTCTGTGGTT TTTGAAAATC ATTGAATTCC 5750
TGCAGCCCGG GGGATCCACT AGTTCTAGTT CTAGAGCGGC CGCTCCAGAA 5800
TTCTAGAAGG TACCCGGGAT CCTTCCTGG GACCCGGCAA GAACCAAAAA 5850
CTCACTCTCT TCAAGGAAAT CCGTAATGTT AAACCCGACA CGATGAAGCT 5900
TGTGTTGGA TGGAAAGGAA AAGAGTTCTA CAGGGAAACT TGGACCCGCT 5950
TCATGGAAGA CAGCTTCCCC ATTGTTAACG ACCAAGAAGT GATGGATGTT 6000
TTCCCTTGTG TCAACATGCG TCCCCTAGA CCCAACCGTT GTTACAAATT 6050
CCTGGCCCAA CACGCTCTGC GTTGCACCC CGACTATGTA CCTCATGACG 6100
TGATTAGGAT CGTCGAGCCT TCATGGGTGG GCAGCAACAA CGAGTACCGC 6150
ATCAGCCTGG CTAAGAAGGG CGGCGGCTGC CCAATAATGA ACCTTCACTC 6200
TGAGTACACC AACTCGTTCG AACAGTTCAT CGATCGTGTC ATCTGGGAGA 6250
ACTTCTACAA GCCCATCGTT TACATCGGT ACGACTCTGC TGAAGAGGAG 6300
GAAATTCTCC TTGAAGTTTC CCTGGTGTTC AAAGTAAAGG AGTTTGCACC 6350
AGACGCACCT CTGTTCACTG GTCCGGCGTA TTAAAACACG ATACATTGTT 6400
ATTAGTACAT TTATTAAGCG CTAGATTCTG TGCCTGTTG ATTTACAGAC 6450
AATTGTTGTA CGTATTTAA TAATTCATTA AATTATAAT CTTTAGGGTG 6500
GTATGTTAGA GCGAAAATCA AATGATTTC AGCGTCTTA TATCTGAATT 6550

TAAATATTAA ATCCTCAATA GATTTGTAAG ATAGGTTCG ATTAGTTCA 6600
AACAAAGGGTT GTTTTCCGA ACCGATGGCT GGACTATCTA ATGGATTTC 6650
GCTCAACGCC ACAAAACTTG CCAAATCTG TAGCAGCAAT CTAGCTTGT 6700
CGATATTCGT TTGTGTTTG TTTTGTAAATA AAGGTTGAC GTCGTTCAA 6750
ATATTATGCG CTTTGTATT TCTTCATCA CTGTCGTTAG TGTACAATTG 6800
ACTCGACGTA AACACGTTAA ATAAAGCTTG GACATATTAA ACATCGGGCG 6850
TGTTAGCTTT ATTAGGCCGA TTATCGTCGT CGTCCCAACC CTCGTCGTTA 6900
GAAGTTGCTT CCGAAGACGA TTTGCCATA GCCACACGAC GCCTATTAAT 6950
TGTGTCGGCT AACACGTCCG CGATCAAATT TGTAGTTGAG CTTTTGGAA 7000
TTATTTCTGA TTGCGGGCGT TTTTGGCGG GTTTCAATCT AACTGTGCC 7050
GATTTAATT CAGACAACAC GTTAGAAAGC GATGGTGCAG GCGGTGGTAA 7100
CATTTCAGAC GGCAAATCTA CTAATGGCGG CGGTGGTGGA GCTGATGATA 7150
AATCTACCAT CGGTGGAGGC GCAGGCAGGG CTGGCGGCGG AGGCGGAGGC 7200
GGAGGTGGTG GCGGTGATGC AGACGGCGGT TTAGGCTCAA ATGTCTCTT 7250
AGGCAACACA GTCGGCACCT CAACTATTGT ACTGGTTCG GGCGCCGTT 7300
TTGGTTTGAC CGGTCTGAGA CGAGTGCGAT TTTTTCGTT TCTAATAGCT 7350
TCCAACAATT GTTGTCTGTC GTCTAAAGGT GCAGCGGGTT GAGGTTCCGT 7400
CGGCATTGGT GGAGCGGGCG GCAATTCAAGA CATCGATGGT GGTGGTGGTG 7450
GTGGAGGCGC TGGAATGTTA GGCACGGGAG AAGGTGGTGG CGGCGGTGCC 7500
GCCGGTATAA TTTGTTCTGG TTTAGTTGT TCGCGCACGA TTGTGGGCAC 7550
CGGCGCAGGC GCCGCTGGCT GCACAAACGGA AGGTCGTCTG CTTCGAGGCA 7600

GCGCTTGGGG TGGTGGCAAT TCAATATTAT AATTGGAATA CAAATCGTAA 7650
AAATCTGCTA TAAGCATTGT AATTCGCTA TCGTTTACCG TGCCGATATT 7700
TAACAACCGC TCAATGTAAG CAATTGTATT GTAAAGAGAT TGTCTCAAGC 7750
TCCGCACGCC GATAACAAGC CTTTCATTT TTACTACAGC ATTGTAGTGG 7800
CGAGACACTT CGCTGTCGTC GACGTACATG TATGCTTGT TGTCAAAAAC 7850
GTCGTTGGCA AGCTTTAAAA TATTAAAAG AACATCTCTG TTCAGCACCA 7900
CTGTGTTGTC GTAAATGTTG TTTTGATAA TTTGCGCTTC CGCAGTATCG 7950
ACACGTTCAA AAAATTGATG CGCATCAATT TTGTTGTTCC TATTATTGAA 8000
TAAATAAGAT TGTACAGATT CATATCTACG ATTGTCATG GCCACCACAA 8050
ATGCTACGCT GCAAACGCTG GTACAATTT ACGAAAACGT CAAAAACGTC 8100
AAAACTCGGT ATAAAATAAT CAACGGCGC TTTGGCAAAA TATCTATTTC 8150
ATCGCACAAG CCCACTAGCA AATTGTATT GCAGAAAACA ATTCGGCGC 8200
ACAATTTAA CGCTGACGAA ATAAAAGTTC ACCAGTTAAT GAGCGACCAC 8250
CCAAATTTA TAAAAATCTA TTTAATCAC GGTTCCATCA ACAACCAAGT 8300
GATCGTGATG GACTACATTG ACTGTCCGA TTTATTTGAA AACTACAAA 8350
TTAAAGGCGA GCTTCGTAC CAACTTGTAA GCAATATTAT TAGACAGCTG 8400
TGTGAAGCGC TCAACGATT GCACAAGCAC AATTCTACAC ACAACGACAT 8450
AAAACTCGAA AATGTCTTAT ATTCGAAGC ACTTGATCGC GTGTATGTTT 8500
GCGATTACGG ATTGTGCAA CACGAAAACG CACTTAGCGT GCACGACGGC 8550
ACGTTGGAGT ATTTTAGTCC GGAAAAAATT CGACACACAA CTATGCACGT 8600
TTCGTTGAC TGGTACGCGG CGTGTAAACA TACAAGTTGC TAACCAGCGG 8650
TTCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA 8700

CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTAA AGCCTGGGT 8750
GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCGC 8800
TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC 8850
GCGCGGGGAG AGGC GGTTTG CGTATTGGGC GCTCTCCGC TTCCTCGCTC 8900
ACTGACTCGC TGCGCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA 8950
CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGAT AACGCAGGAA 9000
AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC 9050
GCGTTGCTGG CGTTTTCCA TAGGCTCCGC CCCCCTGACG AGCATCACAA 9100
AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT 9150
ACCAGGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC 9200
CTGCCGCTTA CCGGATACCT GTCCGCCTT CTCCCTTCGG GAAGCGTGGC 9250
GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTT 9300
GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC 9350
GCCTTATCCG GTAATATCG TCTTGAGTCC AACCCGGTAA GACACGACTT 9400
ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG 9450
TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT 9500
AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG 9550
AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAAACCACC GCTGGTAGCG 9600
GTGGTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT 9650
CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA 9700
AAACTCACGT TAAGGGATT TGTCATGAG ATTATCAAAA AGGATCTTCA 9750

CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA 9800
TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC 9850
TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC TGACTCCCCG 9900
TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG CCCCAGTGCT 9950
GCAATGATAAC CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT 10000
AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT 10050
CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAAGCTAG AGTAAGTAGT 10100
TCGCCAGTTA ATAGTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT 10150
GGTGTACCGC TCGTCGTTTG GTATGGCTTC ATTCAAGCTCC GGTTCCCAAC 10200
GATCAAGGCG AGTTACATGA TCCCCCATGT TGTGCAAAAA AGCGGTTAGC 10250
TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTATAC 10300
ACTCATGGTT ATGGCAGCAC TGCATAATTCTCTTACTGTC ATGCCATCCG 10350
TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA 10400
TAGTGTATGC GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGATAAA 10450
TACCGCGCCA CATAGCAGAA CTTAAAAGT GCTCATCATT GGAAAACGTT 10500
CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG 10550
ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC 10600
CAGCGTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAGG 10650
GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTCAA 10700
TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT 10750
TGAATGTATT TAGAAAAATA AACAAATAGG GGTCGCGC ACATTCCCC 10800
GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC 10850

TATAAAAATA GGCGTATCAC GAGGCCCTT CGTCTCGCGC GTTTCGGTGA 10900
TGACGGTGAA AACCTCTGAC ACATGCAGCT CCCGGAGACG GTCACAGCTT 10950
GTCTGTAAGC GGATGCCGGG AGCAGACAAG CCCGTCAGGG CGCGTCAGCG 11000
GGTGTGGCG GGTGTCGGGG CTGGCTTAAC TATGCGGCAT CAGAGCAGAT 11050
TGTACTGAGA GTGCACCATA TATGCGGTGT GAAATACCGC ACAGATGCGT 11100
AAGGAGAAAA TACCGCATCA GGCGCCATTG GCCATTCAAGG CTGCGCAACT 11150
GTTGGGAAGG GCGATCGGTG CGGGCCTCTT CGCTATTACG CCAGCTGGCG 11200
AAAGGGGGAT GTGCTGCAAG GCGATTAAGT TGGGTAACGC CAGGGTTTTC 11250
CCAGTCACGA CGTTGTAAAA CGACGGCCAG TGCC 11284

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 397 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met	Ala	Lys	Phe	Leu	Val	Asn	Val	Ala	Leu	Leu	Leu	Leu	Leu	Leu
1				5					10					15
Leu	Leu	Ser	Gly	Ala	Trp	Ala	His	Met	Arg	Ser	His	His	His	His
					20				25					30
His	His	Met	Pro	Ala	Gly	Ser	Trp	Asp	Pro	Ala	Gly	Tyr	Leu	Leu
						35			40					45
Tyr	Cys	Pro	Cys	Met	Gly	Arg	Phe	Gly	Asn	Gln	Ala	Asp	His	Phe
							50		55					60
Leu	Gly	Ser	Leu	Ala	Phe	Ala	Lys	Leu	Leu	Asn	Arg	Thr	Leu	Ala
						65			70					75

Val	Pro	Pro	Trp	Ile	Glu	Tyr	Gln	His	His	Lys	Pro	Pro	Phe	Thr
				80						85				90
Asn	Leu	His	Val	Ser	Tyr	Gln	Lys	Tyr	Phe	Lys	Leu	Glu	Pro	Leu
				95						100				105
Gln	Ala	Tyr	His	Arg	Val	Ile	Ser	Leu	Glu	Asp	Phe	Met	Glu	Lys
					110					115				120
Leu	Ala	Pro	Thr	His	Trp	Pro	Pro	Glu	Lys	Arg	Val	Ala	Tyr	Cys
					125					130				135
Phe	Glu	Val	Ala	Ala	Gln	Arg	Ser	Pro	Asp	Lys	Lys	Thr	Cys	Pro
					140					145				150
Met	Lys	Glu	Gly	Asn	Pro	Phe	Gly	Pro	Phe	Trp	Asp	Gln	Phe	His
					155					160				165
Val	Ser	Phe	Asn	Lys	Ser	Glu	Leu	Phe	Thr	Gly	Ile	Ser	Phe	Ser
					170					175				180
Ala	Ser	Tyr	Arg	Glu	Gln	Trp	Ser	Gln	Arg	Phe	Ser	Pro	Lys	Glu
					185					190				195
His	Pro	Val	Leu	Ala	Leu	Pro	Gly	Ala	Pro	Ala	Gln	Phe	Pro	Val
					200					205				210
Leu	Glu	Glu	His	Arg	Pro	Leu	Gln	Lys	Tyr	Met	Val	Trp	Ser	Asp
					215					220				225
Glu	Met	Val	Lys	Thr	Gly	Glu	Ala	Gln	Ile	His	Ala	His	Leu	Val
					230					235				240
Arg	Pro	Tyr	Val	Gly	Ile	His	Leu	Arg	Ile	Gly	Ser	Asp	Trp	Lys
					245					250				255
Asn	Ala	Cys	Ala	Met	Leu	Lys	Asp	Gly	Thr	Ala	Gly	Ser	His	Phe
					260					265				270
Met	Ala	Ser	Pro	Gln	Cys	Val	Gly	Tyr	Ser	Arg	Ser	Thr	Ala	Ala
					275					280				285

Pro Leu Thr Met Thr Met Cys Leu Pro Asp Leu Lys Glu Ile Gln
 290 295 300
 Arg Ala Val Lys Leu Trp Val Arg Ser Leu Asp Ala Gln Ser Val
 305 310 315
 Tyr Val Ala Thr Asp Ser Glu Ser Tyr Val Pro Glu Leu Gln Gln
 320 325 330
 Leu Phe Lys Gly Lys Val Lys Val Val Ser Leu Lys Pro Glu Val
 335 340 345
 Ala Gln Val Asp Leu Tyr Ile Leu Gly Gln Ala Asp His Phe Ile
 350 355 360
 Gly Asn Cys Val Ser Ser Phe Thr Ala Phe Val Lys Arg Glu Arg
 365 370 375
 Asp Leu Gln Gly Arg Pro Ser Ser Phe Phe Gly Met Asp Arg Pro
 380 385 390
 Pro Lys Leu Arg Asp Glu Phe
 395 397

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5009 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GAACCAGGCC GATCACTTCT TGGGCTCTCT GGCATTTGCA AAGCTGCTAA 50
 ACCGTACCTT GGCTGTCCCT CCTTGGATTG AGTACCAGCA TCACAAGCCT 100
 CCTTTCACCA ACCTCCATGT GTCCTACCAAG AAGTACTTCA AGCTGGAGCC 150
 CCTCCAGGCT TACCATCGGG TCATCAGCTT GGAGGATTTC ATGGAGAAGC 200

TGGCACCCAC CCACTGGCCC CCTGAGAAGC GGGTGGCATA CTGCTTGAG 250
GTGGCAGCCC AGCGAAGCCC AGATAAGAAG ACGTGCCCCA TGAAGGAAGG 300
AAACCCCTT GCCCATTCT GGGATCAGTT TCATGTGAGT TTCAACAAAGT 350
CGGAGCTTT TACAGGCATT TCCTCAGTG CTTCCTACAG AGAACAAATGG 400
AGCCAGAGAT TTTCTCCAAA GGAACATCCG GTGCTTGCCC TGCCAGGAGC 450
CCCAGCCCAG TTCCCCGTCC TAGAGGAACA CAGGCCACTA CAGAAAGTACA 500
TGGTATGGTC AGACGAAATG GTGAAGACGG GAGAGGCCA GATTGATGCC 550
CACCTTGTCC GGCCCTATGT GGGCATTTCAT CTGCGCATTG GCTCTGACTG 600
GAAGAACGCC TGTGCCATGC TGAAGGACGG GACTGCAGGC TCGCACTTCA 650
TGGCCTCTCC GCAGTGTGTG GGCTACAGCC GCAGCACAGC GGCCCCCTC 700
ACGATGACTA TGTGCCTGCC TGACCTGAAG GAGATCCAGA GGGCTGTGAA 750
GCTCTGGGTG AGGTCGCTGG ATGCCAGTC GGTCTACGTT GCTACTGATT 800
CCGAGAGTTA TGTGCCTGAG CTCCAACAGC TCTTCAAAGG GAAGGTGAAG 850
GTGGTGAGCC TGAAGCCTGA GGTGGCCCAG GTGACCTGT ACATCCTCGG 900
CCAAGCCGAC CACTTATTG GCAACTGTGT CTCCTCCTTC ACTGCCTTG 950
TGAAGCGGGA GCGGGACCTC CAGGGGAGGC CGTCTTCTTT CTTGGCATG 1000
GACAGGCCCT CTAAGCTGCG GGACGAGTTC TGATTCTGGC CGGAGCACCA 1050
GACCCTCTGA TCCTGGAGGG ACCAGAGTCT GAGCTGGTCC TTCCAGCCAG 1100
GCCTGGCAGC CAGAGGTGCT CCGGGATTGC AAACTCCTCT TCTCACCTGC 1150
CAAAGATGGA GAAGAGTGCC AGGGACCCCT CAAGGAGGGA GACGCTCCAT 1200
ATCCCCAGGGC ATAGGACTTG CAGGTTCTA GGAGCAGGAG CATCTCCCAT 1250
CGCACGTGCT TTCTGCTCTT CTGGGAATTCT CTCACACTGG CAAAGCAGTC 1300

CAGCCTCCGT CTTCTGGTCC ACTCTGCTCT GAGCAGCCTG GGATGCTGAA 1350
CTCTTCAGAG AGATTTTTT ATAGAGAGAT TTCTATAATT TTGATACAAG 1400
GTCATGACTA TCCTAGAACT CTCTGTGGTT TTTGAAAATC ATTGAATTCT 1450
ATTAATGTAG GTACCTAAAG TGACCTTAAC TGAATGTGGA TGAGGCTGGG 1500
GCTGGTGTGG GTCTTTGGC TGCTTTCAA GGTGTCCCCC AATGTGGCCC 1550
TCAAGAGCCA TCCCCACTGC CTGGCCAGAG CCATTGTTGT CCCCTACTTC 1600
CTAGGCCATT TCTGGGGCTT GGGGGATGAA TGCTGTCTG TGCTGTAAAC 1650
ACTATGCAA TGGAAGTTAT CGGTTGTGGT GCTGTGCAGC GCTCTGTGGG 1700
CGACTAAGTG CCACTCACGC AGCATGTTCC TGGCAAGGAG CACATACCAT 1750
CAAGCCACAC TATCATGGTA TTGTTCTCAC AGTCTTTGG TGGTTGATGG 1800
CCACTGCAA CCTGGCACCA TCAGATCTCT TCTGATCTCT TGCCCCAGTG 1850
GGGCCTGGTT GGTAGAATGT TGGCATTGG TTGATATCCA AAGCCTGTT 1900
TCCCAGCCGT CCTCCTGCAG CTGGAGCCTT CAGGCCGTAT TCTCACGAGG 1950
GAACGTTTGC CAAGGCTCTG ACCTCACAGA AGATGCCAG GGCCCAGAAG 2000
CCATCAGAAT TATCAGTGGA GAAGCACCTT TTGACTCTTC CCTTCCAATG 2050
TAATCTCTGC CAACACCAG AGGCTTAAGG TGCTCTAAGT CATGAGTGT 2100
TTGGTCTCAA ATGCTGCAGT TTTAATAATC TGTGACTCCT GAGAGCCAT 2150
GGTTTTTGGA CCTTGTGGTT CTAAAATTCC TTGTCTGACC CCTGTAGATC 2200
TTTCCTTGC CATGTCACCT CCCTTGGCCT TTGATCCTGG AAAGGTGGCA 2250
GAGCCTCCAC TGAGCCAGGC CCAGAGCTCC TTGCAGTGCC TTCTCCTTG 2300
TTTACCTGTG GGAGGAAACA CTTTTTTGT CAGGGGCAGC CTGGTTCAGA 2350

GCTCAGAGGT CACACTGTAT CAAAGATCTC AAACAGCAAA GTCAGCATT 2400
GCTGTATAGA GCTGCCACCC AACTCTAAGC AGGAGAAACT GTACAGAAAG 2450
GGCTTGCTA TTTTCCCTT TTGGGAAAAC AATGAAGTGT TTTAAGTCCT 2500
GGGTGGACTG AGAGATGGTT TGCCTGTCCA GACTTGCTCT CAAGCCTCAT 2550
CCAGAGAAGG AGCTGCAGAT GAGGGAGCCC GTACACTCCC TGCCACCACT 2600
AGGTTGTAAG CCTGTAGCTG GCTGGCTGAT TTCATTTGG AATTCAATTG 2650
CCATCCACAG CCTTACACTA GGCACACACT TTAGAGTCTG GGGCTCCAGT 2700
GGGGCCCGCC TAATTTTTT TCCCCCAAG ACAGGGCCTT GCTCTGTCTC 2750
CCAGGCTGGA GTGCAGTGGC ATGATCATGG CTTACTGCAG CCTTGATCTC 2800
CCAGGCTCAA GCGATCCTTC TGCCTCAGCC TCTCTGGTAG CTGAGACTGC 2850
ATGCCAGCT CCAAATCACC TTGATTCTATA TCAGCAGTAA TAATCAATTG 2900
TGTCTGAAA GAAAGGGCAC CAGAAGTTCT AGCAAAATTC AGTTGTGTTTC 2950
TGTGAGCTAG CACTTTTCC TCTGACCCAA TTTTCTTACC TATAAAATGG 3000
TGATAAAAAC CGACAGGTTG TTCAAAGGCC CAGATCAGCT AAAGCATGTA 3050
TATAAGAGCA CGTTGTAAAC TTGAAAGAGA CAAAGGCACA AATGTGGCTG 3100
TTGATTAATT TGACTGCTTC TCGTTGCTCG TCACCTCCAT GCCAGGCACT 3150
GTGCTTGCTA ATTGCTTAT GGGGCATTC TCTTATTTAT TCCCCAGCCC 3200
TGGGAAATAG GAGCTGTCAT TATCCTTCTC TTTCTGCACA AGGAAAAATT 3250
AATGCCCTGA GAATTGTCAT AATTTCCCA AGGCTGCCA GCTGGTGGTG 3300
TTAAGCCAGA ATTTGACCTC CCAGAGCCAG TTTCCATTAG CTGCCATGCT 3350
CTGCTGCCTC TAATTCACAG AATGCACTTT CTACCCCTGTG TGCCATGGAG 3400
ACCTCCTATG GAAAAATGAT CAGCCACCTT ACCTTCTACT GGGTACCTGC 3450

TGTGAGTCTG CCTATGCCAG AAGGATTAAG GAGGGGAGGT TACCCAAGAA 3500
ACAAAGCCTA CATGCCGCTT ACAGCCCCG TTGGATGGTT GCTCAGTACA 3550
ACAGTCTTGC ATTCAAGCAGG TGTTTGTCA TCACCTACTA TGTGTCAGGC 3600
TCTATGCTAG GTACTGGGA TACAGGAGAG AATCAAGCGT AAAGTCTTG 3650
TTCTCAAGGA ATTCGCATTC TAGAAAGTAG AAGATGTAAT AAATGTACTG 3700
TGGGACATGT TAATAAGTGC TATAAAGAAA TATAAAGGGT TTGGGAGCAA 3750
AAAGAGGGAG TGGATCTATT TTAGATGAGC CCAGGTAAGA CCTCTCTGAA 3800
GAGCTGTCAT GAAGGAGGGA GGGAGCACAT TCCTGGCAGA GAAAACAGCA 3850
CGTGCAAAGG CCCCAGACT GGAGTGTGTT CCTGAAGAGC AGCCAGGAGG 3900
CCAGCATGGC TGGAGAGGCA GGCATAGGCA GGGAACCGAG CAGCAGGTCA 3950
GAGCAGGCGA GCTGACATTG TGCAGCCTGG ACGGCCATGG CAGGAAGCTT 4000
TTAGTTGGAG AGATACAGGA AGCCTCCTAG GGTTCTGAGC AGAAGAGGG 4050
CATGAGCTGA TTCACATTCT GAAGGACCTC TCTAGCTGGC CAGTGCTGAG 4100
GAGGTTGGAG AGAGAAAGGG TGAAAGCAGA GAGACCAGTG CAGGGCTGTT 4150
AACAGGGTTG CAGGCGAGAG ACTGGGGTGC TGGGCTCCCC TAGACTAGGA 4200
CTCCAGTGCC CTCCTCTCCC AAGAGACAAA GGCCATTGCA TTGAAGGAGG 4250
TGGGAAATGA TTAGATTCTG AACATATGTA ATTATTTTC AGTCTTTTC 4300
AAAGATACAA ATATTTACAT AGTTTAATC ATGTAATATA TACAATTAA 4350
TGTCTAGTG TTTTACTTAA TAGTGTATCA TGTTTCCCT GTTGGTATGT 4400
AGCCTGGATA AATGCTCTTA ATTATAAAAA ATTCTGTCGA GGAGTGGTCC 4450
ATAGTTTATT GTTTCCCTAT TATGAGAATT TAGGCCAAGT GTGGTGGCTC 4500

ATGCCTGTAA TCCCAGCACT TTGCGAGGCC GAGGTGGGCA GATCACTTGA 4550
GGTGAGGAGT TCAAGACCAG CCTGGCCAAC ATGGTGAATT ATCTCTACTA 4600
AAAATAACAAA AAAATAATAA TAATAGCCAG GCGTGGTGGC ACATGCCTGT 4650
ATTCCCAGCT GCTTGGGAGG CTGAGGCAGG AGAATGGCTT GAACCTGGGA 4700
GGTGGAGGTT GCAGTGAGCC GAGATGGTGC CACTGCATTC CAGCCTGGC 4750
AACAGAGCGA GACTCCATCT CAAAAAAAAG GAGACTTCAT GTGCCCCCAA 4800
TTTTTCACTA TTGTTATTG AAAAAATATT TTTATTTGTA AGAGTTTTC 4850
TTTATTTAAA ATGTTCATTA ATAAAGTTGT TGGACGGAA GCAAAAAAAA 4900
AAAGTTGTTT AAGATAAATT CCCAGAAGTG AATTTGTTAG ATCAAACACT 4950
TAAAACTTT TGTTATGGAA GAATTCAAAT ATAAATAAAA AATTGTGAGT 5000
AATAAAATG 5009

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 474 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met	Ser	Asn	Tyr	Arg	Tyr	Ser	Lys	Leu	Asn	Glu	Glu	Glu	Ile	Ser
1														15
Leu	Glu	Asp	Met	Pro	Ser	Ser	Ala	Asn	Gln	Ile	Leu	Thr	Arg	Gln
														30
Glu	Gln	Ile	Ile	Gln	Glu	Gln	Asp	Asp	Glu	Leu	Glu	Leu	Val	Gly
														45
Asn	Ser	Val	Arg	Thr	Leu	Arg	Gly	Met	Ser	Ser	Met	Ile	Gly	Asp
														60

Glu	Leu	Asp	Gln	Gln	Ser	Thr	Met	Leu	Asp	Asp	Leu	Gly	Gln	Glu
							65							75
Met	Glu	Tyr	Ser	Glu	Thr	Arg	Leu	Asp	Thr	Ala	Met	Lys	Lys	Met
							80					85		90
Ala	Lys	Leu	Thr	His	Leu	Glu	Asp	Gly	Met	Leu	Leu	Ala	Arg	Arg
							95					100		105
Ile	Val	Gln	Ser	Met	Gln	Asn	Asp	His	Gly	Ala	Leu	Ser	Ser	Pro
							110					115		120
Val	Phe	Pro	Arg	Leu	Cys	Pro	Ser	Gly	Leu	Thr	Thr	Tyr	Val	Pro
							125					130		135
Tyr	Ile	Val	Asp	Phe	Ser	Ser	Leu	Thr	Phe	His	Ile	Phe	Ile	Ile
							140					145		150
Ile	Ile	Ile	Ile	Ile	Asp	Phe	Cys	Ser	Gln	Ser	Gln	Ser	Lys	
							155					160		165
Gly	Arg	Phe	Gly	Asn	Gln	Val	Asp	Gln	Phe	Leu	Gly	Val	Leu	Ala
							170					175		180
Phe	Ala	Lys	Ala	Leu	Asp	Arg	Thr	Leu	Val	Leu	Pro	Asn	Phe	Ile
							185					190		195
Glu	Phe	Lys	His	Pro	Glu	Thr	Lys	Met	Ile	Pro	Phe	Glu	Phe	Leu
							200					205		210
Phe	Gln	Val	Gly	Thr	Val	Ala	Lys	Tyr	Thr	Arg	Val	Val	Thr	Met
							215					220		225
Gln	Glu	Phe	Thr	Lys	Lys	Ile	Met	Pro	Thr	His	Phe	Val	Gly	Thr
							230					235		240
Pro	Arg	Gln	Ala	Ile	Tyr	Asp	Lys	Ser	Ala	Glu	Pro	Gly	Cys	His
							245					250		255
Ser	Lys	Glu	Gly	Asn	Pro	Phe	Gly	Pro	Tyr	Trp	Asp	Gln	Ile	Asp
							260					265		270

Val	Ser	Phe	Val	Gly	Asp	Glu	Tyr	Phe	Gly	Asp	Ile	Pro	Gly	Gly
														285
275														
Phe	Asp	Leu	Asn	Gln	Met	Gly	Ser	Arg	Lys	Lys	Trp	Leu	Glu	Lys
														300
290														
Phe	Pro	Ser	Glu	Glu	Tyr	Pro	Val	Leu	Ala	Phe	Ser	Ser	Ala	Pro
														315
305														
Ala	Pro	Phe	Pro	Ser	Lys	Gly	Lys	Val	Trp	Ser	Ile	Gln	Lys	Tyr
														330
320														
Leu	Arg	Trp	Ser	Ser	Arg	Ile	Thr	Glu	Gln	Ala	Lys	Lys	Phe	Ile
														345
335														
Ser	Ala	Asn	Leu	Ala	Lys	Pro	Phe	Val	Ala	Val	His	Leu	Arg	Asn
														360
350														
Asp	Ala	Asp	Trp	Val	Arg	Val	Cys	Glu	His	Ile	Asp	Thr	Thr	Thr
														375
365														
Asn	Arg	Pro	Leu	Phe	Ala	Ser	Glu	Gln	Cys	Leu	Gly	Glu	Gly	His
														390
380														
His	Leu	Gly	Thr	Leu	Thr	Lys	Glu	Ile	Cys	Ser	Pro	Ser	Lys	Gln
														405
395														
Gln	Ile	Leu	Glu	Gln	Ile	Glu	Ala	His	Arg	Gln	Glu	Pro	Asp	Asp
														420
410														
Met	Tyr	Thr	Ser	Leu	Ala	Ile	Met	Gly	Arg	Ala	Asp	Leu	Phe	Val
														435
425														
Gly	Asn	Cys	Val	Ser	Thr	Phe	Ser	His	Ile	Val	Lys	Arg	Glu	Arg
														450
440														
Asp	His	Ala	Gly	Gln	Ser	Pro	Arg	Pro	Ser	Ala	Phe	Phe	Gly	Ile
														465
455														
Arg	Ala	Val	Lys	Arg	His	Ile	Asp	Leu						
470														
474														

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 61 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met	Pro	Ala	Gly	Ser	Trp	Asp	Pro	Ala	Gly	Tyr	Leu	Leu	Tyr	Cys
1					5						10			15
Pro	Cys	Met	Gly	Arg	Phe	Gly	Asn	Gln	Ala	Asp	His	Phe	Leu	Gly
		20								25			30	
Ser	Leu	Ala	Phe	Ala	Lys	Leu	Leu	Asn	Arg	Thr	Leu	Ala	Val	Pro
				35					40				45	
Pro	Trp	Ile	Glu	Tyr	Gln	His	His	Lys	Pro	Pro	Phe	Thr	Asn	Leu
			50						55				60	
His														
61														

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTTCTTGGGC TCTCTGGCAT TTGCAAAGCT GCTAAACCGT 40

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTCGACGATT TGGCATGGAA CCGACAGGGA GGAACCTAAC 40

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CTTCTTGGGC TCTCTGGCAT TTGCAAAGCT GCTAAACCGT 40

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TCCCTGGGGA GTTCCTCCCT CTGCGAGGTA 30

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Arg	Ser	His	His	His	His	His	His	Met	Pro	Ala	Gly	Ser	Trp	Asp	
1								5					10		15

Pro Ala Gly Tyr Leu Leu Tyr Xaa Pro Xaa Met Gly Arg
20 25 28

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 amino acids
(B) TYPE: Amino Acid
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Thr Val Asp Gly Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly
1 5 10 15

Gly Ser Cys Lys Asp Asp Ile Asn Ser Tyr Glu Cys Trp Cys Pro
20 25 30

Phe Gly Phe Glu Gly Lys Asn Cys Glu Leu Asp Val Thr His His
35 40 45

His His His His Gly Ser Ala
50 52

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1100 base pairs
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ATGCCCGCGG GCTCCTGGGA CCCGGCCGGT TACCTGCTCT ACTGCCCTG 50

CATGGGGCGC TTTGGGAACC AGGCCGATCA CTTCTTGGGC TCTCTGGCAT 100

TTGCAAAGCT GCTAAACCGT ACCTTGGCTG TCCCTCCTTG GATTGAGTAC 150

CAGCATCACA AGCCTCCTTT CACCAACCTC CATGTGTCTT ACCAGAAGTA 200

CTTCAAGCTG GAGCCCCTCC AGGCTTACCA TCGGGTCATC AGCTTGGAGG 250
ATTTCATGGA GAAGCTGGCA CCCACCCACT GGCCCCCTGA GAAGCGGGTG 300
GCATACTGCT TTGAGGTGGC AGCCCAGCGA AGCCCAGATA AGAAGACGTG 350
CCCCATGAAG GAAGGAAACC CCTTTGGCCC ATTCTGGGAT CAGTTCATG 400
TGAGTTCAA CAAGTCGGAG CTTTTTACAG GCATTCCTT CAGTGCTTCC 450
TACAGAGAAC AATGGAGCCA GAGATTTCT CCAAAGGAAC ATCCGGTGCT 500
TGCCCTGCCA GGAGCCCCAG CCCAGTTCCC CGTCCTAGAA GAACACAGGC 550
CACTACAGAA GTACATGGTA TGGTCAGACG AAATGGTGAA GACGGGAGAG 600
GCCAGATTG ATGCCCACCT TGTCCGGCCC TATGTGGCA TTCATCTGCG 650
CATTGGCTCT GACTGGAAGA ACGCCTGTGC CATGCTGAAG GACGGGACTG 700
CAGGCTCGCA CTTCATGGCC TCTCCGCAGT GTGTGGGCTA CAGCCGCAGC 750
ACAGCGGCC CCCTCACGAT GACTATGTGC CTGCCTGACC TGAAGGAGAT 800
CCAGAGGGCT GTGAAGCTCT GGGTGAGGTC GCTGGATGCC CAGTCGGTCT 850
ACGTTGCTAC TGATTCCGAG AGTTATGTGC CTGAGCTCCA ACAGCTCTTC 900
AAAGGGAAGG TGAAGGTGGT GAGCCTGAAG CCTGAGGTGG CCCAGGTCGA 950
CCTGTACATC CTCGGCCAAG CCGACCACCT TATTGGCAAC TGTGTCTCCT 1000
CCTTCACTGC CTTTGTGAAG CGGGAGCGGG ACCTCCAGGG GAGGCCGTCT 1050
TCTTTCTTCG GCATGGACAG GCCCCCTAAG CTGCGGGACG AGTTCTGATT 1100

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 343 amino acids
(B) TYPE: Amino Acid
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Asn Gln Ala Asp His Phe Leu Gly Ser Leu Ala Phe Ala Lys Leu
1 5 10 15

Leu Asn Arg Thr Leu Ala Val Pro Pro Trp Ile Glu Tyr Gln His
20 25 30

His Lys Pro Pro Phe Thr Asn Leu His Val Ser Tyr Gln Lys Tyr
35 40 45

Phe Lys Leu Glu Pro Leu Gln Ala Tyr His Arg Val Ile Ser Leu
50 55 60

Glu Asp Phe Met Glu Lys Leu Ala Pro Thr His Trp Pro Pro Glu
65 70 75

Lys Arg Val Ala Tyr Cys Phe Glu Val Ala Ala Gln Arg Ser Pro
80 85 90

Asp Lys Lys Thr Cys Pro Met Lys Glu Gly Asn Pro Phe Gly Pro
95 100 105

Phe Trp Asp Gln Phe His Val Ser Phe Asn Lys Ser Glu Leu Phe
110 115 120

Thr Gly Ile Ser Phe Ser Ala Ser Tyr Arg Glu Gln Trp Ser Gln
125 130 135

Arg Phe Ser Pro Lys Glu His Pro Val Leu Ala Leu Pro Gly Ala
140 145 150

Pro Ala Gln Phe Pro Val Leu Glu Glu His Arg Pro Leu Gln Lys
155 160 165

Tyr Met Val Trp Ser Asp Glu Met Val Lys Thr Gly Glu Ala Gln
170 175 180

Ile His Ala His Leu Val Arg Pro Tyr Val Gly Ile His Leu Arg
185 190 195

Ile Gly Ser Asp Trp Lys Asn Ala Cys Ala Met Leu Lys Asp Gly
200 205 210

Thr Ala Gly Ser His Phe Met Ala Ser Pro Gln Cys Val Gly Tyr
215 220 225

Ser Arg Ser Thr Ala Ala Pro Leu Thr Met Thr Met Cys Leu Pro
230 235 240

Asp Leu Lys Glu Ile Gln Arg Ala Val Lys Leu Trp Val Arg Ser
245 250 255

Leu Asp Ala Gln Ser Val Tyr Val Ala Thr Asp Ser Glu Ser Tyr
260 265 270

Val Pro Glu Leu Gln Gln Leu Phe Lys Gly Lys Val Lys Val Val
275 280 285

Ser Leu Lys Pro Glu Val Ala Gln Val Asp Leu Tyr Ile Leu Gly
290 295 300

Gln Ala Asp His Phe Ile Gly Asn Cys Val Ser Ser Phe Thr Ala
305 310 315

Phe Val Lys Arg Glu Arg Asp Leu Gln Gly Arg Pro Ser Ser Phe
320 325 330

Phe Gly Met Asp Arg Pro Pro Lys Leu Arg Asp Glu Phe
335 340 343